Generalized Ensemble Simulations for Complex Systems

Bernd A. Berg (berg@hep.fsu.edu)

Department of Physics, Florida State University, Tallahassee FL 32306, USA

Abstract

The most efficient MC weights for the calculation of physical, canonical expectation values are not necessarily those of the canonical ensemble. The use of suitably generalized ensembles can lead to a much faster convergence of the simulation. Although not realized by nature, these ensembles can be implemented on computers. In recent years generalized ensembles have in particular been studied for the simulation of complex systems. For these systems it is typical that conflicting constraints lead to free energy barriers, which fragment the configuration space. Examples of major interest are spin glasses and proteins. In my overview I first comment on the strengths and weaknesses of a few major approaches, multicanonical simulations, transition variable methods, and parallel tempering. Subsequently, two applications are presented: a new analysis of the Parisi overlap distribution for the 3d Edwards-Anderson Ising spin glass and the helix-coil transition of amino-acid homo-oligomers.

1. Introduction

Markov chain Monte Carlo (MC) simulations are an indispensable tool for the study of physical models. To calculate expectation values in the Gibbs canonical ensemble, most simulations use Boltzmann weights. However, in the course of time it has become clear that the Boltzmann factor is not always the most efficient weight to generate the desired configurations of the canonical ensemble. In particular before embarking on a large-scale computer simulation, one of the questions which ought to be addressed is "What are suitable weight factors for the problem at hand?" Nowadays simulations of more general ensembles are employed for an increasing number of applications in physics,

chemistry, structural biology and other areas, for reviews see [1–4].

For instance, the multicanonical (MUCA) method [5–10] calculates in one simulation the canonical ensemble at many temperatures. Its best-established application is calculations of interface tensions at first order phase transitions. There, canonically rare configurations are simply enhanced, the *statical* case. This is similar to the importance sampling idea, which for the calculation of canonical expectation values led from naive sampling to the Markov process sampling with the Boltzmann weights. The additional complication is now that the new weights are *a priori* unknown. The nowadays most important applications of the MUCA method are for complex

systems [7,8]. There, the rational of MUCA simulations is to flatten free energy barriers, such that the *dynamics* of the simulation becomes improved. The enhancement of particular configurations can no longer be explicitly controlled and the issue of optimizing MC algorithms for complex systems is far from being well understood, although some improvements have been quite spectacular.

In the next section we introduce the MUCA method, using as illustration the $2d$ ten-state Potts model with its strong first-order phase transition. Subsection 2.1 discusses the weight recursions of Ref.[1,9] and [10]. Random walk and transition matrix methods [11–18] generalize the MUCA approach and are summarized in section 3. Distinct from these methods are multiple Markov chain algorithms [19–21] (parallel tempering). They are introduced in section 4, where their recent combinations [22] with MUCA methods are also sketched. In section 5 we turn to applications for complex systems. Instead of being exhaustive (see the reviews [1–3]) we focus on two recent studies: a new analysis of the Parisi overlap distribution for the 3d Edwards-Anderson Ising spin glass [23] and the helix-coil transition of the polyalanine protein [24,26]. Outlook and conclusions are given in section 6.

2. Multicanonical Simulations

Multicanonical simulations are best established for investigations of first-order phase transitions. For them pseudocritical points $\beta^c(L)$, where L is the lattice size, exist such that the canonical energy density $P(E) = P(E; L)$ becomes double-peaked and the maxima at $E^1_{\max} < E^2_{\max}$ are of equal height $P_{\max} = P(E^1_{\max}) = P(E^2_{\max})$. In between these values a minimum is located at some energy $E_{\min}$. Configurations at $E_{\min}$ are exponentially suppressed like

$$P_{\min} = P(E_{\min}) = c_f\, L^p\, \exp(-f^s A) \qquad (1)$$

where f^s is the interface tension and A is the minimal area between the phases, $A = 2L^{d-1}$ for an L^d lattice, c_f and p are constants. To determine the interface tension with Binder's histogram method, one has to calculate the quantities

$$f^s(L) = -\frac{\ln R(L)}{A(L)} \ \text{ with } \ R(L) = \frac{P_{\min}(L)}{P_{\max}(L)} \qquad (2)$$

and to make a finite size scaling (FSS) extrapolation of $f^s(L)$ for $L \to \infty$. However, for large systems a canonical MC simulation will practically never visit configurations at energy $E = E_{\min}$ and estimates of the ratio $R(L)$ will be very inaccurate. The terminology *supercritical slowing down* was coined to characterize thus an exponential deterioration of simulation results with lattice size. The MUCA method overcomes this problem by sampling, in an appropriate energy range, with an *approximation*

$$\widehat{w}_{mu}(k) = w_{mu}(E^{(k)}) = e^{-b(E^{(k)})\, E^{(k)} + a(E^{(k)})} \qquad (3)$$

to the inverse density of states $1/n(E^{(k)})$. Here the function $b(E)$ is the inverse *microcanonical temperature* at energy E and $a(E)$ is the dimensionless, microcanonical free energy. One samples instead of the canonical energy distribution $P(E)$ a new MUCA distribution

$$P_{mu}(E) = c_{mu}\, n(E)\, w_{mu}(E) \approx c_{mu} \ . \qquad (4)$$

The desired canonical distribution is obtained by re-weighting, which is rigorous, because the weights $w_{mu}(E)$ used in the actual simulation are exactly known. With the approximate relation (4) the average number of configurations sampled does no longer depend strongly on the energy and accurate estimates of the ratio (2) $R(L) = P_{\min}/P_{\max}$ become possible.

The MUCA method requires two steps
(i) Obtain a *working estimate* of the weights (3). Working estimate means that the approximation to the inverse density of states has to be good enough to ensure movement in the desired energy range.
(ii) Perform a Markov chain MC simulation with the weights $\widehat{w}_{mu}(k)$. Canonical expectation values are found by re-weighting to the Gibbs ensemble and standard jackknife methods allow reliable error estimates.

For the $2d$ 10-state Potts model figure 1 reproduces a thus obtained MUCA histogram to-
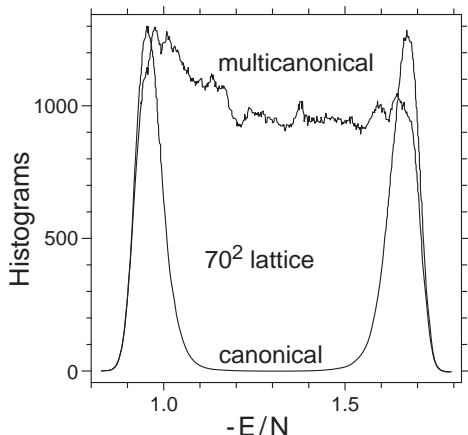
Fig. 1. The multicanonical energy histogram $P_{mu}(E)$ together with its canonically re-weighted energy density $P(E)$ for the 2d 10-state Potts model on a 70×70 lattice [6].

gether with its canonically re-weighted energy histogram [6]. The canonical energy density of this figure gives the estimate $2f^s(L)$ of the interface tension on the 70×70 lattice. Combining the such obtained estimates from several lattices allows the infinite volume FSS extrapolation $2f^s = 0.0978(8)$. At the time of its publication this number was in strong disagreement with other numerical results. Afterwards the exact value was discovered to be $2f^s = 0.094701...$ [27]. This success helped the acceptance of the method, which by now has been applied to a large number of models [1].

There exist many variants of the MUCA method. MUCA refers to calculations of canonical expectation values for a temperature range and re-weighting has to be done in the internal energy. Similarly, other physical quantities can be considered, *e.g. multimagnetical* [28] refers to simulations which give results for a certain range of the magnetic field. A variant for cluster updates due to Janke and Kappler [29] is called *multi-bondic*. In Ref.[30] the *multi-overlap* algorithm was introduced, which focuses on the Parisi order parameter of spin glasses (see section 5). Combining MUCA with multigrid methods has been explored by Janke and Sauer [31]. For molecular dynamics, Langevin and hybrid MC variants see Hansmann et al. [32] and, with emphasise on lattice gauge theory, Arnold et al. [33].

2.1. *Multicanonical Recursion*

For spin systems with first-order phase transitions the FSS behavior is relatively well known. Provided the steps between system sizes are not too large, it is then possible to obtain working estimates of the $w_{mu}(E)$ weights by means of a FSS extrapolation from the already simulated smaller systems [6]. However, this method cannot be applied to complex systems and to get a working estimate of the MUCA parameters by means of a recursion is then at the heart of the method.

As the overall normalization is irrelevant for updating purposes, it is sufficient to consider nearest neighbor (in energy) ratios of the weights (4)

$$R^n(E) = \frac{w(E)}{w(E+\epsilon)} = e^{\epsilon b^n(E)} \qquad (5)$$

where ϵ is the smallest available energy step and n refers to the nth iteration. Initially, $w^0(E) = 1$ is a good starting value, because the system can then move freely. The final equation for the recursion, as derived in [1], reads

$$R^{n+1}(E) = R^n(E) \left[\frac{\hat{H}^n(E+\epsilon)}{\hat{H}^n(E)} \right]^{\hat{g}_0^n(E)} \qquad (6)$$

where the exponent $\hat{g}_0^n(E)$ is recursively determined by the equations

$$\hat{g}_0^n(E) = \frac{g_0^n(E)}{g^n(E) + g_0^n(E)} \ , \qquad (7)$$

$$g_0^n(E) = \frac{H^n(E+\epsilon) \, H^n(E)}{H^n(E+\epsilon) + H^n(E)} \quad \text{and} \qquad (8)$$

$$g^{n+1}(E) = g^n(E) + g_0^n(E), \ g^0(E) = 0 \,. \qquad (9)$$

Typically, we want to cover an energy range

$$E_{\max} - E_{\min} \sim V$$

The optimum for a flat energy distribution is given by a random walk in the energy. This implies a CPU time increase $\sim V^2$ to keep the number of $E_{\max} \to E_{\min} \to E_{\max}$ transitions constant. The recursion (6) needs an additional $\sim V^{0.5}$ (optimum) number of attempts to cover the entire range. For the two-dimensional Ising model on $L \times L$ lattices, with $E_{\min} = -2L^2$ (the ground

3

Table 1

Sweeps per $E_{\max} = 0 \to E_{\min} = -L^2 \to E_{\max} = 0$ transition for $L \times L$ Ising models. Here τ_1 refers to the first transition during the recursion (6), τ_{rec} to the subsequent transitions during the recursion and $\tau_{\mathrm{p}rod}$ to the transitions during the production part.

L	τ_1	τ_{rec}	$\tau_{\mathrm{p}rod}$
10	1654 (66)	576 (14)	545 (14)
20	19573 (730)	3872 (120)	3855 (110)
30	79743 (2600)	12820 (350)	11428 (300)
40	192274 (6900)	25901 (820)	27898 (1700)
50	472103 (13000)	48227 (1600)	45803 (2300)
60	810042 (23000)	78174 (2300)	78288 (7200)
80	2917722 (56000)	178456 (8900)	184236 (9100)

state) and $E_{\max} = 0$ (the average energy at $\beta = 0$), table 1 gives the number of sweeps per transition. The first transition is tedious to find, additional transitions are much faster.

A new, possibly more efficient, recursion was recently proposed by F. Wang and Landau [10] (my interpretation of the merits of the method differs somewhat from the authors). They perform updates with estimators $n(E)$ of the density of states

$$p(E_1 \to E_2) = \min\left[\frac{n(E_1)}{n(E_2)}, 1\right] . \qquad (10)$$

Each time an energy level is visited, the estimator is updated according to

$$n(E) \to n(E)\, f \qquad (11)$$

where, initially, $n(E) = 1$ and $f = f_0 = e^1$. Once the desired energy range is covered, the factor f is refined,

$$f_1 = \sqrt{f},\ f_{n+1} = \sqrt{f_{n+1}} , \qquad (12)$$

until some small value like $f = e^{-8} = 1.00000001$ is reached. A fast convergence towards a good estimator of the spectral density is reported. Afterwards the usual, MUCA production runs may be carried out. A comparison of the recursions (6) and (11,12) is subtle, because in neither case is the optimal schedule or termination point known.

3. Transition Matrix Methods

The basic idea of the approach [11–18] is to employ general transition matrix elements instead of just weight factors. An example is Wang's [17] random walk algorithm, which uses the transition probabilities [11,13–15]

$$p(k \to k') = \min\left(1, \frac{N(E + \triangle E, -\triangle E)}{N(E, \triangle E)}\right) \qquad (13)$$

where $N(E, \triangle E)$ is the microcanonical average for the number of transitions from the configuration k with energy E to the configuration k' with energy $E' = E + \triangle E$. A flat histogram is obtained and the configuration dependence is reduced to energy dependence. The density of states is obtained from the equation [13–15]

$$n(E)\, N(E, \triangle E) = \qquad (14)$$

$$n(E + \triangle E)\, N(E + \triangle E, -\triangle E) .$$

The history and general merits of the approach are reviewed in [4]. There, it is claimed that, in some situations, equation (14) works more efficiently than the Wang-Landau recursion (11,12). However, the issue may not be settled. In particular, it seems [17] that in the random walk MC the use of estimators for $N(E, \triangle E)$ faces more serious problems than the use of estimators for the weights in MUCA simulations. After defining reasonable schedules and termination points, one could compare the transition times of table 1 with those achieved with Wang's random walk approach.

4. Parallel Tempering

The developments sketched above have to be distinguished from the method of multiple Markov chains (*parallel tempering*), which was introduced by Geyer [19] and, independently, by Hukusima and Nemoto [20]. The latter work was influenced by the *simulated tempering* [34] method, which in turn can be understood as a special case of the *method of expanded ensembles* [35]. Parallel tempering is particularly well suited for parallel processing on clusters of workstations.

Parallel tempering performs n canonical MC simulations at different β-values with Boltzmann weight factors

$$w_{B,i}(E^{(k)}) = e^{-\beta_i E^{(k)}}, \ i = 1, \dots, n \qquad (15)$$

with $\beta_1 < \beta_2 < \dots < \beta_{n-1} < \beta_n$ and allows exchange of neighboring β-values

$$\beta_{i-1} \longleftrightarrow \beta_i \quad \text{for} \ i = 2, \dots, n \ . \qquad (16)$$

These transitions lead to the energy change

$$
\begin{aligned}
\triangle E &= \left(-\beta_{i-1} E_i^{(k)} - \beta_i E_{i-1}^{(k')}\right) \qquad (17)\\
&\quad - \left(-\beta_i E_i^{(k)} - \beta_{i-1} E_{i-1}^{(k')}\right)\\
&= (\beta_i - \beta_{i-1}) \left(E_i^{(k)} - E_{i-1}^{(k')}\right)
\end{aligned}
$$

which is accepted or rejected according to the Metropolis algorithm. The β_i-values have to be determined such that a reasonably large acceptance rate is obtained for the β exchange (16) and Ref.[20] employs a recursive method.

Remark: The method works for dynamical but not for statical supercritical slowing down, because each member of the discrete set of weight factors samples still a Boltzmann distribution (*e.g.* it is not a valid method for calculating interfacial tensions). This can possibly be overcome by applying the parallel tempering idea to Gaussian distributions instead of the canonical ensemble (communicated to the author by T. Neuhaus).

Ref.[22] explores earlier introduced [36] combinations of parallel tempering, called replica-exchange method in [22], with MUCA methods. The replica-exchange MUCA algorithm performs a short replica-exchange simulation to determine the MUCA weight factors and continues with the production part of a regular MUCA simulation, whereas MUCA replica-exchange works with replicas of the MUCA ensemble in different energy ranges.

5. Applications to Complex Systems

The relevance of MUCA methods for complex systems was pointed out early [7]. In these systems one encounters large free energy barriers due to disorder and frustration. Multicanonical simulations try to overcome the barriers through excursions into the disordered phase. In addition, one may try to identify and enhance relevant rare configurations. Here, we limit our interest to two recent studies of (a) spin glasses and (b) proteins. For reviews of applications to complex systems see [1–3].

5.1. *Spin glasses*

The Parisi overlap parameter q is defined as the overlap of two replicas of the system at identical temperatures

$$q = \sum_{i=1}^{N} s_i^1 s_i^2 \ . \qquad (18)$$

The free-energy structure of the spin glass system is thought to be intimately related to the barriers in the Parisi order parameter distributions. The *multi-overlap* variant [30] of the MUCA method enhances the minima of the Parisi overlap parameter densities. Using the method in large-scale simulations led to a variety of new insights, for instance about the self-averaging properties of free energy barriers [37]. The enhancement is most dramatic for the tails of the averaged Parisi order parameter distribution. Recently, this allowed us to demonstrate a conjectured relation between the overlap probability density and extreme order statistics over about 80 orders of magnitude [23], see figure 2.

5.2. *Proteins*

Proteins are linear polymers with the 20 naturally occurring amino acids as monomers. Chains smaller than a few tens of amino acids are called peptides. The problem is to predict the folded conformation of proteins and peptides solely from their amino acid sequence. For many years the emphasis of numerical investigations has been on finding the global minimum potential energy and the major difficulty encountered is the multiple minima problem. Molecular dynamics has been the numerical method of first choice, but the fraction of stochastic investigations shows an increasing trend.
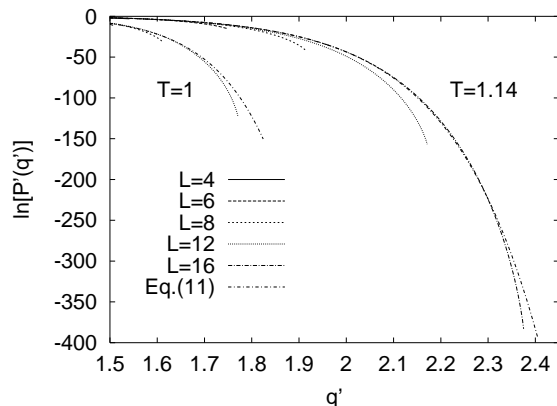
Fig. 2. Tails of the rescaled overlap (18) probability density in comparison with predictions of extreme order statistics [Eq.(11)], see [23] ($T = 1.14$ is at the freezing temperature).

The major advantage of MUCA and related methods in the context of proteins is that they allow for investigations of the *thermodynamics* of the entire free energy landscape of the protein. This was realized by Hansmann and Okamoto [8] when they introduced MUCA sampling to the problem of protein folding and, slightly later, by Hao and Scheraga [38]. Since then numerous applications have been performed and the simulations have been quite successful for peptides. By now a quite extensive literature exists, which is compiled in [2,3]. In Ref.[39] the weight recursion (6) is adapted to these systems with continuous energy.

A particularly nice application is the helix formation of polyalanine by means of a MUCA simulation [24,25], depicted in figure 3. Up to $N = 30$ amino acids are studied numerically and a phase transition develops at $T_c = (541 \pm 8)\,K$ in the infinite volume limit (pseudocritical temperature for finite systems are in the range from $427\,K$ ($N = 10$) to $518\,K$ ($N = 30$)). FSS estimates of the critical indices α, ν and γ are consistent with a first-order transition. Progress is made concerning the inclusion of aqueous solutions into the simulation [26].
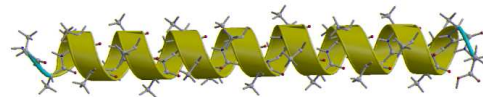

Fig. 3. A helix configuration from a multicanonical simulation of polyalanine [24] (courtesy Ulrich Hansmann and Yuko Okamoto).

6. Outlook and Conclusions

Sampling of broad energy distributions allows supercritical slowing down to be overcome. This is well established for first-order phase transitions. Systems with conflicting constraints remain, despite some progress, notoriously difficult and for them most hope lies in achieving further algorithmic improvements. Reasonably large peptides, at the border to proteins, are promising systems for a complete coverage. Transition matrix extensions of the MUCA method and its combinations with parallel tempering appear promising.

References

[1] B.A Berg, Fields Institute Communications **26** (2000) 1 [cond-mat/9909236].

[2] U.H. Hansmann and Y. Okamoto, Ann. Rev. Comp. Phys. **6** (1999) 129.

[3] A. Mitsutake, Y. Sugita and Y. Okamoto, Biopolymers (Peptide Science) **60** (2001) 96.

[4] J.-S. Wang and R.H. Swendsen, cond-mat/0104418.

[5] G.M. Torrie and J.P. Valleau, J. Comp. Phys. **23** (1977) 187.

[6] B.A. Berg and T. Neuhaus, Phys. Rev. Lett. **68** (1992) 9.

[7] B.A. Berg and T. Celik, Phys. Rev. Lett. **69** (1992) 2292.

[8] U.H. Hansmann and Y. Okamoto, J. Comp. Chem. **14** (1993) 1333.

[9] B.A. Berg, J. Stat. Phys. **82** (1996) 323.

[10] F. Wang and D.P. Landau, Phys. Rev. Lett. **86** (2001) 2050.

[11] B.A. Berg, Nature **61** (1993) 708.

[12] G.R. Smith and A.D. Bruce, J. Phys. A **28** (1995) 6623; Phys. Rev. E **53** (1996) 6530.

[13] P.M.C. de Oliveira, T.J.P. Penna and H.J. Herrmann, Braz. J. Phys. **26** (1996) 677.

[14] B.A. Berg and U.H.E. Hansmann, Eur. Phys. J. B **6** (1998) 396.

[15] P.M.C. de Oliveira, Eur. Phys. J. B **6** (1998) 111.

[16] J.-S. Wang, T.K. Tay and R.H. Swendsen, Phys. Rev. Lett. **82** (1999) 476.

[17] J.-S. Wang, Eur. Phys. J. B **8** (1999) 287. J.-S. Wang and L.W. Lee, Comp. Phys. Commun. **127** (2000) 131.

[18] J.D. Muñoz and H.J. Herrmann, Int. J. Mod. Phys. C **10** (1999) 95; Comp. Phys. Commun. **121/2** (1999) 13.

[19] C.J. Geyer, in *Computing Science and Statistics*, Proceedings of the 23rd Symposium on the Interface, E.M. Keramidas (editor), Interface Foundation, Fairfax, Virginia, 1991, pp.156–163.

[20] K. Hukusima and K. Nemoto, J. Phys. Soc. Japan **65** (1996) 1604.

[21] M.C. Tesi, E.J. Janse van Rensburg, E. Orlandini and S.G. Whittington, J. Stat. Phys. **82** (1996) 155.

[22] Y. Sugita and Y. Okamoto, Chem. Lett. **329** (2000) 261.

[23] B.A. Berg, A. Billoire and W. Janke, Phys. Rev. E **65** (2002) 045102R.

[24] U.H. Hansmann and Y. Okamoto, J. Chem. Phys. **110** (1999) 1267; *ibid.* **111** (1999) 1339 (E).

[25] N.A. Alves and U.H. Hansmann, Phys. Rev. Lett. **84** (2000) 1836.

[26] A. Mitsutake and Y. Okamoto, J. Chem. Phys. 112 (2000) 10638

[27] C. Borgs and W. Janke, J. de Physique I **2** (1992) 2011.

[28] B.A. Berg, U.H. Hansmann and T. Neuhaus, Phys. Rev. B **47** (1993) 497

[29] W. Janke and S. Kappler, Phys. Rev. Lett. **74** (1995) 212.

[30] B.A. Berg and W. Janke, Phys. Rev. Lett. **80** (1998) 4771.

[31] W. Janke and T. Sauer, Phys. Rev. E **49** (1994) 3475.

[32] U.H. Hansmann, Y. Okamoto and E. Eisenmenger, Chem. Phys. Lett. **259** (1996) 321.

[33] G. Arnold, Th. Lippert and K. Schilling, Phys. Rev. D **59** (1999), 054509.

[34] E. Marinari and G. Parisi, Europhys. Lett. **19** (1992) 451.

[35] A.P. Lyubartsev, A.A. Martsinovski, S.V. Shevkanov and P.N. Vorontsov-Velyaminov, J. Chem. Phys. **96** (1992) 1776.

[36] U.H. Hansmann, Chem. Phys. Letters **281** (1997) 140

[37] B.A. Berg, A. Billoire and W. Janke, Phys. Rev. B **61** (2000) 12143.

[38] M.-H. Hao and H.A. Scheraga, J. Phys. Chem. **98** (1994) 4940.

[39] F. Yasar, T. Celik, B. Berg and H. Meirovitch, J. Comp. Chem. **21** (2000) 1251.